Bayesowskie testowanie modeli tobitowych w analizie spłaty kredytów detalicznych

Wstęp

Podstawowym narzędziem wspomagającym pracę analityka bankowego przy ocenie wniosku kredytowego jest system scoringowy. Jądrem każdego systemu scoringowego jest mechanizm, który umożliwia, na podstawie tylko wybranych, ale najistotniejszych cech klienta, predykcję ryzyka związanego z udzieleniem mu kredytu. Jest to model klasyfikacji kredytobiorców do danych grup ryzyka. W uproszczeniu można powiedzieć, że jego celem jest kalkulacja, dla konkretnego wniosku kredytowego, wskaźnika informującego o poziomie ryzyka kredytowego (stopniu wypłacalności klienta), w oparciu o mierzalne i niemierzalne cechy klienta. Na tej podstawie klient – potencjalny kredytobiorca – jest kwalifikowany najczęściej do jednej z dwóch rozłącznych grup, tj. grupy klientów podwyższonego bądź obniżonego ryzyka, złych i dobrych. Bank uzależnia swoją decyzję o udzieleniu kredytu od tego, do której grupy ryzyka klient został przydzielony. Osoby z pierwszej grupy nie kwalifikują się do uzyskania kredytu ze względu na wysokie ryzyko kredytowe. Udzielenie kredytu osobom z drugiej grupy wiąże się z akceptowalnym poziomem ryzyka.

W celu klasyfikacji kredytobiorców stosuje się najczęściej modele statystyczno-matematyczne. Najbardziej znanymi metodami statystycznymi wykorzystywanymi przy konstrukcji modeli scoringowych są m.in. analiza dyskryminacyjna, drzewa klasyfikacyjne, metoda najbliższego sąsiedztwa oraz modele danych jakościowych. W literaturze polskiej, o zastosowaniu wspomnianych metod w systemach scoringowych pisano m.in. w pracach [2], [3], [5], [6] i [7].

Spośród modeli danych jakościowych najczęściej stosuje się modele z dychotomiczną lub wielomianową zmienną endogeniczną: model probitowy lub logitowy. Natomiast modele tobitowe (regresji cenzurowanej), których konstrukcja jest zbliżona do modeli danych jakościowych, są rzadko spotykane w systemach scoringowych. Zastosowanie modelu tobitowego pozwala na uwzględnienie pełniejszej informacji o zmiennej endogenicznej, której wartości są cenzurowane, ale mierzone na skali mocniejszej – ilorazowej.

Głównym celem pracy jest prezentacja bayesowskiego porównywania mocy wyjaśniającej czterech różnych specyfikacji modelu tobitowego, które zastosowano w empirycznej analizie opóźnienia w spłacie rat kapitałowo-odsetkowych w przypadku kredytów detalicznych polskiego banku.

Zastosowane proste uogólnienie standardowego modelu tobitowego, zaproponowanego przez Tobina w 1958 r., polega na: (i) przyjęciu rozkładu z szerszej klasy niż normalny, tj. rozkładu *t* Studenta o nieznanej liczbie stopni swobody, (ii) wprowadzeniu zależności między zmienną ukrytą z_t a zmiennymi egzogenicznymi w formie wielomianu stopnia drugiego, zamiast zależności liniowej. Fakt, że (i) własności metody największej wiarygodności nie zostały w pełni poznane w przypadku regresji z rozkładem *t* Studenta oraz że (ii) dla modelu tobitowego brak jest „dobrych" miar dobroci dopasowania modelu do danych, były motywacją do zastosowania wnioskowania bayesowskiego.

1. Model tobitowy z rozkładem *t* Studenta.

Rozważmy uogólnienie standardowego modelu tobitowy postaci

$$\begin{cases} y_t = \max(z_t, 0) \quad \text{dla} \quad t = 1, \ldots, T, \\ z_t = G(w_t, \beta) + \varepsilon_t, \end{cases} \tag{1}$$

gdzie składnik losowy ε_t ma jednowymiarowy rozkład *t* Studenta z ν stopniami swobody, z parametrem niecentralności, czyli modalną μ równą zero i precyzją τ ($\nu>0$ i $\tau>0$). Funkcję gęstości rozkładu zmiennej ε_t zapisujemy w formie $f_S(\varepsilon_t | \mu = 0, \nu, \tau)$, natomiast funkcja $G(w_t, \beta)$ jest zdefiniowana następująco

$$G(w_t, \beta) = x_t \cdot \beta = \beta_1 + \sum_h \beta_h w_{th} + \sum_h \sum_{i \geq h} \beta_{hi} w_{th} w_{ti} \tag{2}$$

Powyższy, najogólniejszy model będzie określany, jako model II rzędu, zob. [11], w odróżnieniu do modelu I rzędu, w którym zakłada się, że zmienna ukryta z_t jest liniową funkcją zmiennych objaśniających w_{th}, czyli że $\beta_{hi} = 0$ dla każdego h oraz $i \geq h$. Z punktu widzenia estymacji, modele I i II rzędu różnią się jedynie liczbą parametrów, czyli wymiarem wektora β. W obu przypadkach z_t jest liniową funkcją β. Jedną z korzyści z wprowadzenia aproksymacji II rzędu jest to, że w modelu (1) efekty krańcowe względem ustalonej zmiennej w_{th} mogą posiadać przeciwne znaki dla różnych obserwacji, natomiast w modelu I rzędu posiadają identyczne znaki, zob. [9]. Ponadto konstrukcja modelu (1) może przyczynić się do lepszego dopasowania modelu do danych, gdyż wielomian stopnia drugiego (względem w_{th}) jest lepszą aproksymacją (wyższego rzędu) niż wielomian stopnia pierwszego. Natomiast wprowadzenie dodatkowego, kluczowego parametru ν umożliwia naturalne uwzględnienie standardowego modelu tobitowego jako przypadku granicznego, gdy $\nu \to +\infty$ i $\beta_{hi} = 0$.

2. Specyfikacja modelu bayesowskiego

W modelu tobitowym z rozkładem *t* Studenta uogólniona funkcja gęstości *T*-wymiarowego wektora niezależnych obserwacji $y = (y_1, \ldots, y_T)'$, czyli rozkład próbkowy jest iloczynem dystrybuanty i gęstości rozkładu *t* Studenta

$$p(y|\theta) = \prod_{t:y_t=0} F_S(-x_t\beta|\theta) \cdot \prod_{t:y_t>0} f_S(y_t|x_t\beta,\nu,\tau). \tag{3}$$

gdzie $F_S(a)$ dystrybuantą rozkładu *t* Studenta zmiennej ε_t, obliczoną w punkcie a. Bayesowski model statystyczny jest jednoznacznie zdefiniowany poprzez łączny rozkład prawdopodobieństwa dla obserwacji i parametrów $p(y,\theta) = p(y|\theta) \cdot p(\theta)$, gdzie $p(\theta)$ to tzw. brzegowy rozkład parametrów, zwany rozkładem a priori. Podejście bayesowskie wymaga dodatkowo określenia rozkładu a priori na przestrzeni nieznanych parametrów $\theta = (\beta'\ \nu\ \tau)' \in R^k \times R^+ \times R^+$. Dla parametru precyzji, który jest wspólnym parametrem w rozważanych, szczegółowych modelach tobitowych, przyjęto niewłaściwy rozkład a priori postaci $p(\tau) \propto \tau^{-1}$. W przypadku pozostałych parametrów przyjmujemy właściwe rozkłady a priori, gdyż to umożliwia bayesowskie testowanie rozważanych w pracy modeli. Zastosowanie właściwego rozkład a priori dla β i ν pozwala porównywać odpowiednio model I rzędu i II rzędu oraz model z rozkładem normalnym i z rozkładem *t* Studenta. Ponadto, właściwy rozkład a priori dla ν gwarantuje istnienie rozkładu a posteriori dla tego parametru; zob. [4]. Dla β przyjęto a priori rozkład normalny scentrowany wokół wektora zerowego ($\beta^* = 0$), o diagonalnej macierzy kowariancji postaci $H^{*-1} = s_*^2 \cdot I_k$, gdzie s_*^2 jest wariancją a priori dla pojedynczej składowej wektora β. Dla s_*^2 przyjęto wartość 1000, gdyż wówczas implikowane rozkłady a priori dla wielkości podlegających bezpośredniej interpretacji, tj. efektów krańcowych, są mocno rozproszony (z grubymi ogonami). Mniejsze wartości dla s_*^2, np. $s_*^2 = 10$, czyniły te rozkłady silnie skupionymi wokół zera. Natomiast dla parametru ν, który jest parametrem swoistym w modelu z rozkładem *t* Studenta, przyjęto standardowo rozkład wykładniczy o wartości oczekiwanej i odchyleniu standardowym równym $r = 10$ lub $r = 40$, zob. także [4], [9] i [11]. Dobór stałej r nie miał wpływu na wyniki a posteriori dotyczące parametru ν.

Następnie, korzystając ze wzoru Bayesa, gęstość łącznego rozkładu a posteriori można zapisać w formie

$$p(\nu,\beta,\tau|y) = \frac{1}{p(y)} p(\nu,\beta,\tau) \cdot \prod_{t:y_t=0} F_S(-x_t\beta|\nu,\beta,\tau) \cdot \prod_{t:y_t>0} f_S(y_t|x_t\beta,\nu,\tau), \tag{4}$$

gdzie $p(y)$ jest brzegową gęstością wektora obserwacji, zaś

$$p(\nu,\beta,\tau) \propto f_N^{(k)}(\beta|\beta^*, s_*^2 I_k) \cdot f_{EXP}(\nu|r) \cdot \tau^{-1} \tag{5}$$

Skomplikowana struktura jądra rozkładu (4) nie pozwala na analityczne wyznaczenie postaci rozkładów a posteriori bądź podstawowych momentów tego rozkładu, zarówno dla oryginalnych parametrów, jak i innych interesujących wielkości, np. wartości oczekiwanej zmiennej y_t czy efektów krańcowych. W tym przypadku wykorzystano, jak w pracach [9] i [11], algorytm Metropolisa i Hastingsa, jedną z metod Monte Carlo opartych na łańcuchach Markowa.

3. Porównywanie mocy wyjaśniającej modeli

Bayesowskie porównywanie parami wykluczających się modeli, sprowadza się do wyboru takiego modelu, który charakteryzuje się największym prawdopodobieństwem a posteriori, które liczone jest wg wzoru Bayesa:

$$p(M_i|\,y)=\frac{p(y|\,M_i)\cdot p(M_i)}{\sum_{j=1}^{J} p(y|\,M_j)\cdot p(M_j)} \quad \text{dla } i\in\{1,\ldots,J\}, \tag{6}$$

gdzie $p(y|\,M_i)$ i $p(M_i)$ są odpowiednio brzegową gęstością wektora obserwacji y i określonym przez badacza prawdopodobieństwem a priori modelu M_i. Narzędziem, które umożliwia oddzielenie subiektywnego przekonania badacza o prawdziwości sformułowanych hipotez od niezależnej informacji, zawartej na ten temat w danych empirycznych, jest czynnik Bayesa. Reprezentuje on ciężar dowodów dostarczanych przez dane, a świadczących na korzyść jednego z modeli. Dla pary modeli liczymy go wg formuły (zob. [4])

$$BF_{ij} = p(y|M_i)\big/ p(y|M_j). \tag{7}$$

4. Wyniki empiryczne

Przedmiotem badania jest zbiór 39034 kredytów konsumpcyjnych i hipotecznych, udzielonych klientom detalicznym przez bank komercyjny w okresie 01.01.2000 - 30.09.2001 r. Zmienna endogeniczna reprezentuje opóźnienie ze spłatą rat kapitałowo–odsetkowych przez kredytobiorców, jakie zaobserwowano na dzień 30.09.2002 r. Wielkość opóźnienia, wyrażonego w dniach, definiujemy jako różnicę między datą 30.09.2002 r. a ustaloną w harmonogramie spłaty kredytu datą ostatniej niespłaconej w całości raty kapitałowo–odsetkowej. Opóźnienie wynosi zero, gdy kredyt spłacany jest w terminie przewidzianym w umowie lub gdy został już całkowicie spłacony. Zbiór potencjalnych zmiennych egzogenicznych wyjaśniających ryzyko pojedynczej umowy kredytowej zawierał (jak we wcześniejszych pracach autora): płeć, wiek kredytobiorcy, wpływy na rachunki typu ROR, posiadanie rachunku ROR, informację o tym, czy kredytobiorca posiada karty płatnicze lub kredytowe wydane przez ten bank, sposób udzielenia kredytu (przez pośrednika kredytowego albo bezpośrednio przez bank), typ kredytu (kredyt konsumpcyjny albo hipoteczny), okres trwania umowy kredytowej, kwota przyznanego kredytu, waluta kredytu, podstawowe źródło dochodu uzyskiwanego przez kredytobiorcę (umowa o pracę, albo renta lub emerytura, albo własna działalność, umowa o dzieło lub umowa zlecenie, albo inne źródło). Szczegółowe informacje na ten temat znajdują się w pracach [9], [10] lub [11].

W pracy rozważyliśmy cztery specyfikacje ($J = 4$), model najogólniejszy z rozkładem *t* Studenta o nieznanej liczbie stopni swobody (ν) i z częścią regresyjną w postaci wielomianu II stopnia (M_1), następnie model I rzędu z rozkładem *t* Studenta (M_2), model z rozkładem normalnym i ze specyfikacją II rzędu (M_3) oraz standardowy model tobitowy (M_4). Rozkłady próbkowe modeli M_2, M_3 i M_4 uzyskujemy z modelu M_1 poprzez warunkowanie względem wektora β lub parametru ν. Podobnie postępujemy konstruując rozkłady a priori.

Zaproponowano dwa kierunki rozszerzenia modelu standardowego. Pierwszy polega na wprowadzeniu rozkładu o grubych ogonach – model M_2, drugi na uwzględnieniu większej liczby czynników wyjaśniających zaobserwowane wartości zmiennej y_t. Rodzi to następujące pytania: czy w świetle posiadanego materiału empirycznego dwukierunkowe uogólnienie modelu M_4 w formie modelu M_1 jest potrzebne? Czy wystarczyłoby tylko jedno z tych uogólnień, jeżeli tak, to które? Zatem, który z modeli lepiej opisuje badane zjawisko?

Odpowiedzi na te pytania uzyskano obliczając czynniki Bayesa i prawdopodobieństwa a posteriori dla poszczególnych modeli. Tabela 1 przedstawia szczegółowe rezultaty, w zależności od przyjętych prawdopodobieństw a priori dla każdej ze specyfikacji. Rozważano wybór najlepszego modelu, w pierwszym przypadku przyjmując identyczne prawdopodobieństwa a priori dla każdego modelu, zaś w drugim faworyzując specyfikacje oszczędnie sparametryzowane, tj. model standardowy (M_4) i model t Studenta z aproksymacją liniową dla z_t (M_2). Wartości czynników Bayesa wskazują, iż dane empiryczne tak zdecydowanie faworyzują model M_1, że założenie badacza dotyczące prawdopodobieństw a priori dla każdego modelu nie wpływa na wybór najlepszego. Wyniki a posteriori wskazują, że najbardziej prawdopodobnym jest model M_1, który charakteryzuje się największą mocą wyjaśniającą i jest kilkadziesiąt rzędów wielkości bardziej prawdopodobny od drugiego w kolejności modelu M_3. Prawdopodobieństwo a posteriori pozostałych modeli, w tym także M_3, jest bliskie zero. Natomiast, spośród dwóch możliwych rozszerzeń standardowego modelu M_4, dane liczbowe zdecydowanie wskazują na aproksymację II rzędu (model M_3). Natomiast, jeżeli uzależni się prawdopodobieństwo a priori modelu od jego prostoty, to model M_2 jest o osiem rzędów bardziej prawdopodobny a posteriori niż model M_3. Zatem, samo zastosowanie aproksymacji kwadratowej, poprzez zwiększenia liczby parametrów β z 14 do 79, nie wystarcza, aby model M_3 był najlepszy. Podobnie, zastosowanie rozkładu z grubymi ogonami. Dopiero połączenie tych dwóch koncepcji spowodowało, że otrzymano najlepszy model M_1.

Tabela 1.

Brzegowe gęstości wektora obserwacji i prawdopodobieństwa a posteriori badanych modeli.

Model	M_1	M_2	M_3	M_4
Liczba parametrów (k')	81	16	80	15
$\ln p(y \mid M_i)$	–84413	–84602	–84577	–84745
BF	1	$8{,}4\times10^{-83}$	$4{,}2\times10^{-72}$	$8{,}4\times10^{-145}$
Log_{10} BF	–	–82	–71	–144
$p(M_i)$ $p(M_i \mid y)$	0,25 1	0,25 $8{,}4\times10^{-83}$	0,25 $4{,}2\times10^{-72}$	0,25 $8{,}4\times10^{-145}$
$p(M_i) \propto 2^{-k_i'}$ $p(M_i \mid y)$	9×10^{-21} 1	0,3333 $3{,}1\times10^{-63}$	$3{,}6\times10^{-20}$ $1{,}7\times10^{-71}$	0,6666 $1{,}2\times10^{-124}$

Źródło: opracowanie własne.

W modelu M_1 wartość oczekiwana a posteriori dla kluczowego parametru ν wynosi 9,675, przy odchyleniu standardowym 0,622, co potwierdza uzyskane wcześniej wyniki, że prawdopodobieństwo a posteriori modelu z rozkładem normalnym wynosi praktycznie zero.

Tabela 2 prezentuje spodziewane opóźnienie spłaty kredytu dla przeciętnego kredytobiorcy. Model standardowy w stosunku do pozostałych modeli daje najniższe oczekiwane opóźnienie ze spłatą kredytu. Na podstawie modelu *t* Studenta (M_1) wnioskujemy, iż spodziewane przeciętne opóźnienie w spłacie kredytu, bez względu na przebieg spłaty, wynosi 89 dnia, czyli prawie 3 miesiące. Natomiast, jeżeli klient miałby problemy ze spłatę rat kapitałowo-odsetkowych, to spodziewamy się, iż opóźnienie wyniesie 236 dni, czyli ponad 7 miesięcy. Niepewność związana z tymi wielkościami, wyrażona poprzez odchylenie standardowe a posteriori jest niewielka i wynosi od 1 do 9 dni. Przypomnijmy, że zgodnie z uchwałami Komisji Nadzoru Bankowego, gdy opóźnienie w spłacie rat kapitałowo-odsetkowych kredytu konsumpcyjnego przekracza 6 miesięcy, to bank zobowiązany jest do tworzenia rezerw celowych w kwocie równej wartości kredytu pozostającego do spłaty (pomniejszonej o ewentualne zabezpieczenia). Zatem z punktu widzenia zarządzania ryzykiem kredytowym w banku, interesującym zagadnieniem jest określenie siły i kierunku oddziaływania czynników egzogenicznych wpływających na kształtowanie się wielkości opóźnienia spłaty kredytu.

Tabela 2.

Wartości oczekiwane i odchylenia standardowe a posteriori dla średniej wartości $E(y_t \mid y_t > 0)$ i $E(y_t)$.

Model	M_1		M_2		M_3		M_4	
	E(|y)	D(|y)	E(|y)	D(|y)	E(|y)	D(|y)	E(|y)	D(|y)
$E(y \mid y > 0)$	236	9	219	2	224	8	205	2
$E(y)$	89	8	72	1	93	7	72	1

Źródło: obliczenia własne

Spośród zmiennych objaśniających znaczący wpływ na okres zwłoki w spłacie kredytu posiada sposób udzielenia kredytu oraz informacja o tym, czy klient posiadana rachunek ROR w badanym banku. Udzielenie kredytu poprzez pośrednika powoduje zwiększenie hipotetycznego opóźnienia o ponad 117 dni (prawie 4 miesiące) oraz wydłużenie już istniejącego opóźnienia (y_t>0) o 76 dni. Posiadanie rachunku ROR zmniejsza opóźnienie przeciętnie o 28 dni albo zmniejsza je o 18 dni, gdy kredytobiorca nie spłaca regularnie kredytu. Wzrost wieku i wpływów klienta oraz krótki okres trwania kredytu powoduje zmniejszenie opóźnienia w spłacie, aczkolwiek z punku widzenia zarządzania ryzykiem wpływ ten jest mało znaczący. Ponadto, udzielenie kredytu klientom prowadzącym własną działalność gospodarczą powoduje niewielki wzrost opóźnienia ze spłatą, w stosunku do osób zatrudnionych na umowę o pracę, aczkolwiek wnioskowanie o tych wielkościach jest obarczone niepewnością, gdyż charakteryzują się one dużymi wartościami odchyleń standardowych a posteriori. Natomiast pobieranie renty i emerytura lub stypendium zmniejsza potencjalne zadłużenie o 4 lub 8 dni, gdy kredytobiorca nie spłaca kredytu. Te ostatnie źródła dochodu stwarzają minimalne zagrożenie związane z ryzykiem kredytowym, mniejsze nawet niż umowa o pracę. Szczegółowe informacje o efektach krańcowych i oczekiwanym opóźnieniu ze spłatą kredytu, także w przypadku wybranych kredytobiorców, są prezentowane w pracach [9] i [10].

Podsumowanie

Wyniki a posteriori wskazują, że spośród czterech specyfikacji, model *t* Studenta o prawie dziesięciu stopniach swobody i z aproksymacją kwadratową charakteryzuje się największą mocą wyjaśniającą, najlepszym dopasowaniem. W warstwie empirycznej uzyskano interesujące wyniki dotyczące spodziewanego opóźnienia ze spłatą kredytu w przypadku przeciętnego kredytobiorcy. Możliwość prognozowania opóźnienia, a tym samym określenia ewentualnych kosztów utraconych korzyści (związanych z rezerwami celowymi), daje przewagę modelom tobitowym nad modelami dychotomicznej zmiennej endogenicznej. Wydaje się, że modele tobitowe i podejście bayesowskie są interesującymi kierunkami dalszego rozwoju modeli i metod statystycznych, stosowanych w scoringu bankowym.

Literatura

1. Amemiya T., *Tobit Models: A Survey*, „Journal of Econometrics" 1984, 24.
2. Gruszczyński M., *Modele i prognozy zmiennych jakościowych w finansach i bankowości*, Monografie i Opracowania SGH, nr 6, Warszawa 2001.
3. Janc A., M. Kraska, *Credit – scoring. Nowoczesna metoda oceny zdolności kredytowej*. Biblioteka Menadżera i Bankowca, Warszawa 2001.
4. Koop G. *Bayesian Econometrics*, Wiley, Chichester 2003.
5. Kulawik J., *Modele scoringowe w kredytowaniu rolnictwa USA i Kanady*, „Bank i Kredyt" 1996, 27, nr 7-8.
6. Kuryłek W., *Credit scoring – podejście statystyczne*, „Bank i Kredyt" 2000, nr 6.
7. Lasek M., *Data Mining. Zastosowanie w analizach i ocenach klientów bankowych*. Biblioteka Menadżera i Bankowca, Oficyna Wydawnicza „Zarządzanie i Finanse" Sp. z o.o., Warszawa 2002.
8. Maddala G., *Limited-dependent and Qualitative Variables in Econometrics*, Cambridge University Press, Cambridge 1983.
9. Marzec J., *Bayesowski model tobitowy z rozkładem t Studenta w analizie niespłacalności kredytów*, [w:] *Metody ilościowe w naukach ekonomicznych*, red. A. Welfe, Wydawnictwo SGH, Warszawa 2005.
10. Marzec J., *Zastosowanie standardowego modelu tobitowego w prognozowaniu spłacalności kredytów*, [w:] *Prognozowanie w zarządzaniu firmą*, red. P. Dittmann, Prace Naukowe AE, Wrocław 2005, w druku
11. Osiewalski J., J. Marzec, *Model dwumianowy II rzędu i skośny rozkład Studenta w analizie ryzyka kredytowego*, „Folia Oeconomica Cracoviensia" 2004, vol. 45.

Streszczenie

Głównym celem badań było przedstawienie bayesowskiego testowania konkurencyjnych specyfikacji w odniesieniu do modeli tobitowych. Modele te zastosowano w empirycznej analizie niespłacalności kredytów detalicznych.

Przyjęto, iż w modelu tobitowym zmienna objaśniana reprezentuje okres (w dniach) opóźnienia w spłacie kredytu detalicznego (rat i odsetek). W przypadku znaczącej liczby kredytobiorców zaobserwowano terminową spłatę zadłużenia. Wówczas zmienna objaśniana przyjmuje wartość zero, zaś w pozostałych przypadkach przyjmuje wartości większe od zera.

Wykorzystano dwa uogólnienia najczęściej stosowanego modelu regresji cenzurowanej, tj. standardowego modelu tobitowego. Pierwsze uogólnienie polegało na zastosowaniu rozkładu *t* Studenta o nieznanej liczbie stopni swobody jako naturalnego uogólnienia rozkładu normalnego. Drugie rozszerzenie opierało się na przyjęciu aproksymacji kwadratowej zamiast liniowej dla ciągłej zmiennej ukrytej, determinującej zaobserwowane wartości zmiennej cenzurowanej. Jednakże przyjęcie rozkładu *t* Studenta wymagało zastosowania innej metody estymacji niż metoda największej wiarygodności.

Podejście bayesowskie zostało realizowane poprzez zastosowanie metod Monte Carlo opartych na łańcuchach Markowa (ang. *Markov Chain Monte Carlo*). Spośród rozważanych specyfikacji dokonano wyboru najlepszego modelu w oparciu o prawdopodobieństwa a posteriori każdego z nich. Wyniki a posteriori wskazały, iż specyfikacje opierające się na rozkładzie *t* Studenta i aproksymacji kwadratowej (zamiast liniowej) lepiej opisują badane zjawisko niż modele prostsze (np. z rozkładem normalnym).

Bayesian comparison of Tobit models for the analysis of the consumer loans repayment

Summary

The main point of this research is to present Bayesian comparison of various Tobit models which are using in analysis of the consumer loans repayment. A explained variable is expressed as follows: a delay (in days) in paying off loan. Than the dependent variable is zero for a significant fraction of the observations, but in the other cases, are great than zero.

Extensions of the classical Tobit model are used. The first generalization relies on it, that the errors are Student-t distributed with unknown degrees of freedom. Secondly, in the standard tobit model, a latent variable is assumed to represent a utility associated with loans repayment is linear combination of the explanatory variables. We applied second-order polynomial. The uses t-Student distribution requires employment other method of estimation than maximum likelihood estimator. In Bayesian approach we use an Markov Chain Monte Carlo methods as the most feasible numerical methods for the estimation. Bayes factors were then used to compare the evidence one model against competitive models. The data give very strong evidence that Student-t model with second-order approximation compared to the other models.